Entropy in Classical and Quantum Information Theory

William Fedus
Physics Department, University of California, San Diego.

Entropy is a central concept in both classical and quantum information theory, measuring the uncertainty and the information content in the state of a physical system. This paper reviews classical information theory and then proceeds to generalizations into quantum information theory. Both Shannon and Von Neumann entropy are discussed, making the connection to compressibility of a message stream and the generalization of compressibility in a quantum system. Finally, the paper considers the application of Von Neumann entropy in entanglement of formation for both pure and mixed bipartite quantum states.

CLASSICAL INFORMATION THEORY

In statistical mechanics, entropy is the logarithm of the number of arrangements a system can be configured and still remain consistent with the thermodyanmic observables. From this original formulation, entropy has grown to become an important element in many diverse fields of study. One of the first examples was in 1948 when Claude Shannon adopted entropy as a measure of the uncertainty in a random variable, or equivalently, the expected value of information content within a message. Classical information theory, as established by Claude Shannon, sought to resolve two central issues in signal processing

1. The compression achievable for a message while preserving the fidelity of the original information
2. The rate at which a message can be communicated reliabily over a noisy channel

both of these address the issue of the degree of redundancy embedded within a message and, as we will see, entropy is found to be a rigorous and useful approach in this context.

Shannon Entropy and Data Compression

As alluded to previously, classical information theory is framed in terms of the information content and the transmission of messages. A message may be defined as a string of letters chosen from an alphabet of k letters $\{a_1, a_2, ..., a_k\}$. Assuming that each letter a_x occurs with probability $p(a_x)$ and is independently transmitted, this may be denoted as an ensemble $A = \{a_x, p(a_x)\}$; the Shannon entropy of an ensemble is then defined as

$$H(A) \equiv H\left(p(a_x), ..., p(a_k)\right) \equiv -\sum_x p(a_x)\text{log}p(a_x). \quad (1)$$

where log is taken to be base-2 since we are transmitting messages with binary *bits*. This formula is important as it can be used to quantify the resources necessary to store information. For instance, if we use a block code which assigns integers to typical sequences, the information in a string of n letters can be compressed to $H(A)$ bits.

With this framework and definition, we may now consider the maximum compression of a length n message without loss of information. The number of bits necessary to transmit the message is given by

$$H(A^n) = nH(A). \quad (2)$$

which simply states that one needs (n times the entropy of the ensemeble A)-bits.

However, a more interesting case may be considered if we allow a small error $\delta > 0$ in the encoding of the message. Specifically, we will only encode messages in set $B \subset A$ such that $P(B) \geq 1 - \delta$. The information size is given by $H_\delta(A)$ which is equal to $H(A)$ in the limit that $\delta \to 0$. Shannon's noiseless coding theorem then states that

$$\lim_{n\to\infty} \frac{1}{n} H_\delta(A^n) = H(A) \quad (3)$$

that is, the optimal compression rate is simply the Shannon entropy; this is Shannon's noiseless coding theorem.

As a basic example of this theorem, consider an information source which produces messages composed of the four element alphabet,

$$A = \{a_1, a_2, a_3, a_4\}. \quad (4)$$

Clearly, without any compression, two bits of storage are necessary for each letter transmitted (bit strings 00, 01, 10, 11) as depicted in the left side of Figure 1. However, suppose now that symbol a_1 is produced with probability 1/2, symbol a_2 with probability 1/4 and symbols a_3 and a_4 with probability 1/8. We find that by applying Eq. 1, we can calculate the entropy to be,

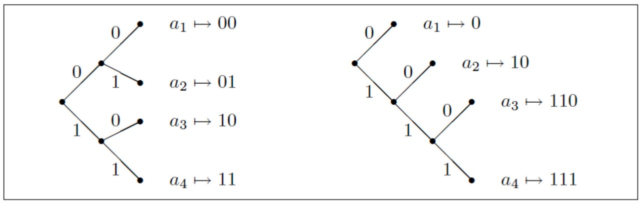

FIG. 1. Tree representation for the encoding of the above example.

$$H(A) = \sum_x -p(a_x)\log p(a_x) \tag{5a}$$

$$= -\frac{1}{2} * \log\left(\frac{1}{2}\right) - \frac{1}{4}\log\left(\frac{1}{4}\right) - 2 * \frac{1}{8}\log\left(\frac{1}{8}\right) \tag{5b}$$

$$= \frac{7}{4} \tag{5c}$$

Therefore, a coding scheme exists in which the messages may be transmitted with *only 7/4 bits* per letter on average if we assume that the message being transmitted is of length n and $n \gg 1$. To realize this compression, we choose to encode a_1 as it string 0, a_2 as bit string 10, a_3 as bit string 110, and a_4 as bit string 111, we find that the *average* length of the compressed string is

$$\frac{1}{2} * 1 + \frac{1}{4} * 2 + \frac{1}{8} * 3 + \frac{1}{8} * 3 = \frac{7}{4} \tag{6}$$

indicating that we have achieved the best possible compression for an encoding. Also, it should be noted that this encoding will not compress *all* messages since clearly an unlucky and improbable message of repeated a_4s would result in a higher average bit content per letter than both encodings. Also, note that Figure 1 explicitly demonstrates that our encoding must not have duplicate prefixes, otherwise, the encoding is not valid. If letters of the alphabet share a common prefix, message strings may not be uniquely decodable.

QUANTUM INFORMATION THEORY

In classical information theory, we have considered a message of n letters, where $n \gg 1$ and each letter is drawn independently from an ensemble $A = \{a_x, p(a_x)\}$. The information content of the ensemble is equal to the Shannon information, $H(A)$, in the limit that $n \to \infty$

To generalize this formulation to the quantum regime, the message may still be regarded as composed of n letters, but now the letters are drawn from an ensemble of quantum states, $X = \{p_x, \rho_x\}$ where ρ_x is the quantum state and p_x is the *a priori* probability of selecting that quantum state. This quantum state may be completely characterized via the density matrix

$$\rho = \sum_x p_x \rho_x \tag{7}$$

The Von Neumann entropy of a quantum state ρ is then defined as

$$S(\rho) \equiv -\mathrm{Tr}\left(\rho \log \rho\right) \tag{8}$$

where log is taken to be base d, d being the dimension of the Hilbert space containing ρ.

Von Neumann entropy enters quantum information theory in three important ways. First, it is a measure of the *quantum* information content of letters in the ensemble, specifically, how many *quibits* are needed to encode the message without loss of information. Second, it is also a measure of the *classical* information content per letter, where we find the maximum amount of information in bits that can be recovered from the message. And finally, as we will see later, Von Neumann entropy is used to quantify the entanglement of pure and mixed bipartite quantum states [1].

Now, returning to Eq. 8, if λ_xs are the eigenvalues that diagonalize ρ, then we may rewrite it as,

$$S(\rho) = -\sum_x \lambda_x \log \lambda_x. \tag{9}$$

One now observes that in this orthonormal basis, the Von Neumann entropy is equivalent to the Shannon entropy, Eq. 1,

$$S(\rho) = H(A) \tag{10}$$

for the ensemble $A = \{a, \lambda_a\}$. This indicates that if a quantum system is a pure separable system, it reduces to the classical system. For a separable quantum system, the Von Neumann entropy is another quantification of the incompressibility of the information content, analagous to the Shannon entropy of a classical system.

Connection to Thermodynamics

Entropy in statistical thermodynamics and in information theory are not disjoint concepts. For instance, if we take an open quantum system (in contact with its environment), we may use the mathematical properties of subadditivity and entropic invariance under unitary evolution for Von Neumann entropy to imply the second law of thermodynamics.

To see this, subadditivity may be succinctly expressed as,

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B) \tag{11}$$

where ρ_A is the partial trace over B ($\rho_A \equiv \mathrm{Tr}_B(\rho_{AB})$) and ρ_B is defined similarly. If we consider the system (A) and the environment (E) to be initially uncorrelated, we can decompose the density matrix as the tensor product $\rho_{AB} = \rho_A \otimes \rho_B$ which by Eq. 11 implies that the entropy is simply the sum of the two states A and B

$$S(\rho_{AB}) = S(\rho_A) + S(\rho_B). \tag{12}$$

If we now allow the system to evolve under a unitary operator U which acts on the full system, the density matrix becomes

$$\rho'_{AE} = U\rho_{AE}U^{-1} \tag{13}$$

but we know that the entropy of a system is invariant under unitary evolution $S(\rho) = S(U\rho U^{-1})$ so therefore,

$$S(\rho'_{AE}) = S(\rho_{AE}). \tag{14}$$

Finally, by invoking Eq. 11 again on the state ρ'_{AE} and using Eq. 12 we find that

$$S(\rho'_A) + S(\rho'_B) \geq S(\rho_{AE}) \tag{15a}$$
$$S(\rho'_A) + S(\rho'_B) \geq S(\rho_A) + S(\rho_B) \tag{15b}$$

so therefore, the basic mathematical properties of Von Neumann entropy implies that the entropy of the universe, that is, the sum of A and E, may not *decrease*, analagous to the second law of thermodynamics.

Quantum Data Compression

Now, as in the classical case, Von Neumann entropy provides us a tool to address the issue of redundancy within a message. However, now in the quantum regime, we instead are seeking to compress the message into a smaller Hilbert space, $\mathcal{H}$, without compromising the fidelity of the message.

As before, consider a message of length n with each letter drawn from an ensemble of pure states $\{p_x, |\psi_x\rangle\}$, where each $|\psi_x\rangle$ is not necessarily orthonormal. We may characterize each letter via the density matrix

$$\rho = \sum_x p_x \,|\psi_x\rangle \langle\psi_x| \tag{16}$$

and the entire message of length n can then be characterized as the n-tensor product

$$\rho^n = \rho \otimes \cdots \otimes \rho \tag{17}$$

Now, the best optimal compression of this quantum message to a Hilbert space $\mathcal{H}$ as $n \to \infty$ is given by

$$\log(dim\mathcal{H}) = S(\rho^n) = nS(\rho) \tag{18}$$

where dim is the dimension of the Hilbert space. This implies that Von Neumann entropy is equal to the number of quibits of quantum information carried for each letter of the message Note that compression is always possible provided that the density matrix is not a maximially mixed state, $\rho = \frac{1}{2}\mathbb{1}$, since we cannot compress random quibits, directly analagous to the classical incompressibility of random bits.

ENTANGLEMENT MEASURES

Entanglement is a feature of quantum states to exhibit correlations that cannot be accounted for classically. Entanglement theory is underpinned by the paradigm of Local Operations and Classical Communication (LOCC) formulated by Bennett *et al* [2]. Under this approach, a quantum state exists between different parties who can perform arbitrary local unitary operations and may communicate with each other over a classical channel. A state is *separable* if it can be created via only local operations and classical communication, otherwise, the state is considered *entangled*.

Mathematically, a pure bipartite state ψ_{AB} is defined to be *entangled* if the following tensor decomposition is not possible

$$|\psi_{AB}\rangle = |\psi_A\rangle \otimes |\psi_A\rangle \tag{19}$$

and similarly, a mixed bipartite state ρ_{AB} is defined to be *entangled* if the density matrix may not be written as

$$\rho_{AB} = \rho_A \otimes \rho_B \tag{20}$$

Since entangled states cannot be generated locally, it may be considered as a resource of the system. This resource is important in different tasks such as quantum computation, quantum communication and quantum cryptography. [3]

Many important bipartite entanglement measurements exist (Schmidt number, k-concurrences, I-concurrence, Negativity, Convex roof extended negativity (CREN), etc.), however, in this paper, we shall consider an entropy-based entanglement measure, *entanglement of* *formation* [3]. We will first consider a *pure* quantum state and then we will consider the more complicated example of a *mixed* quantum state.

Entanglement of Formation of Pure States

For any bipartite system in a pure state, Bennet *et al* [2] demonstrated that it is reasonable to define the entanglement of the system as the von Neumann entropy of either of its two parts. Given a density matrix ρ_{AB} of a pair of quantum systems A and B, consider all possible pure-state decompositions according to Eq. 7.

For each pure state, the entanglement, E, may then be defined as the entropy of either quantum subsystem, A or B

$$E(\psi) = -\mathrm{Tr}\rho_A \mathrm{log}\rho_A = -\mathrm{Tr}\rho_B \mathrm{log}\rho_B. \tag{21}$$

This definition, utilizing the Von Neumann entropy, indicates the information content of the reduced state. Note that this entanglement measure, E, ranges from 0 for a separable pure state to $\mathrm{log}N$ for a maximally entangled state.

Entanglement of Formation of Mixed States

To extend this definition to *mixed* states, the entanglement of formation is defined as the minimum of the average entanglement of each possible pure state of the decomposition,

$$E(\rho) = \min \sum_i p_i E(\psi_i) \tag{22}$$

The probability weighted sum of each individual entanglement, $E(\psi_i)$ is not unique since an infinite number of decompositions exist, therefore, we take the minimum. We choose the minimum (as opposed to another property) since a state which has *any* decomposition as a sum of separable pure states should have 0 entanglement and therefore the density matrix should be found to have 0 entanglement.

For a pair of quibits, Eq. 22 can be written as an explicit function of ρ. To rewrite this equation as such, one must first introduce the spin flip transformation. For a pure state of a single quibit, the spin flip transformation, denoted by a tilde, is defined as

$$\left|\tilde{\psi}\right\rangle = \sigma_y \left|\psi^*\right\rangle \tag{23}$$

where $|\psi^*\rangle$ is the complex conjugate of $|\psi\rangle$ in a fixed basis such as $\{|\uparrow\rangle, |\downarrow\rangle\}$ and σ_y in the same basis is the standard Pauli matrix. For n quibits, the spin flip operation may be applied to each individual quibit and for a density matrix, each σ_y is applied to both the right and left sides. For instance,

$$\tilde{\rho} = (\sigma_y \otimes \sigma_y)\, \rho^* \,(\sigma_y \otimes \sigma_y) \tag{24}$$

where the complex conjugate is in a fixed basis $\{|\uparrow\uparrow\rangle, |\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle, |\downarrow\downarrow\rangle\}$

With this definition, Hill and Wootters [4] demonstrated that Eq. 22 can written as,

$$E(\psi) = \mathcal{E}(C(\psi)) \tag{25}$$

where the concurrence C is defined as

$$C(\psi) \equiv |\left\langle \psi|\tilde{\psi} \right\rangle| \tag{26}$$

and $\mathcal{E}$ is defined as

$$\begin{aligned}\mathcal{E}(C) \equiv &-\frac{1+\sqrt{1-C^2}}{2}\mathrm{log}\frac{1+\sqrt{1-C^2}}{2} \\ &-\frac{1-\sqrt{1-C^2}}{2}\mathrm{log}\frac{1-\sqrt{1-C^2}}{2}\end{aligned} \tag{27}$$

This form of entanglement can be motivated by the fact that it allows us to transfer our analysis to the determination of the concurrence. We observe that $\mathcal{E}$ monotonically increases from $0 \to 1$ as the concurrence increases over that same range, therefore, concurrence may also be regarded as a valid measure of entanglement. This proves useful as we may extend the definition to that of arbitrary bipartite *mixed* states.

With the spin flip and the function $\mathcal{E}(C)$, we can now write the entanglement of formation of a *mixed* state ρ as

$$E(\rho) = \mathcal{E}(C(\rho)) \tag{28}$$

where

$$C(\rho) = \max\{0, \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4\} \tag{29}$$

where the λ_is are the eigenvalues, in decreasing order, of the Hermitian matrix $R \equiv \sqrt{\sqrt{\rho}\tilde{\rho}\sqrt{\rho}}$ and each λ_i is a non-negative real number. Eq. 28 was proved by Wootters [5] to hold for arbitrary states of a two quibit system. For brevity, the proof will be omitted, but the usefulness of entropy in entanglement measures is well-demonstrated.

[1] J. Preskill, "Lecture notes on quantum computation," (1998).

[2] C. Bennett, D. DiVincenzo, J. Smolin, and W. Wootters, Physical Review A **54**, 3824 (1996).
[3] C. Eltschka and J. Siewert, , 8 (2014), arXiv:1402.6710.
[4] S. Hill and W. Wootters, Physical Review Letters **78**, 5022 (1997).
[5] W. Wootters, Physical Review Letters **14**, 199 (1998).